

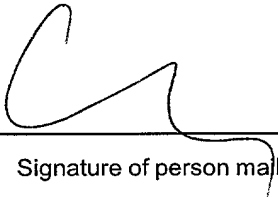
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Certificate under 37 CFR 1.10 of Mailing by "Express Mail"

ET 154872247 US
"Express Mail" label number

9/13/01
Date of Deposit

I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.



Signature of person mailing correspondence

Cary Tope-McKay

Typed or printed name of person mailing correspondence

Note: Each paper must have its own certificate of mailing by "Express Mail".

A METHOD AND APPARATUS FOR AUTOMATICALLY EXTRACTING METADATA FROM ELECTRONIC DOCUMENTS USING SPATIAL RULES

TECHNICAL FIELD

5 The present invention relates generally to the extraction of metadata from electronic documents. More specifically, this invention relates to a combination of text based matching and spatial reasoning used in the extraction of metadata.

BACKGROUND

10 Digital libraries have been introduced to the Internet and are utilized to store a variety of documents and provide retrieval services for the documents. Documents in digital libraries include journal articles, conference papers, technical reports, and dissertations. Most digital libraries retrieve relevant documents utilizing a keyword-based search in human-generated database indices. Some systems automatically generate citation indices
15 from a document, providing a framework for literature retrieval by following citation links. Evaluation of the document is based on the number of citations, and identification of research trends. The above-described system locates, downloads, and parses certain electronic files to extract citations from the documents in order to produce the citation index. However, this system does not extract other useful information from the
20 document such as title, author, and affiliations.

A fundamental step in automatically introducing electronic documents into a digital library system is to disaggregate each document into its basic constituents, so a reader

can effectively index, search, and disseminate the document. For example, in a scientific paper, metadata such as authors, affiliations, title, abstract, and citations play a fundamental role in consolidating the knowledge of the reader. Therefore, it is important to extract such metadata in an efficient and accurate manner.

5

In the past, various systems have been presented to disaggregate text-based documents. They generally fall into one of the following two general categories. The first category is, context-free grammar parsing. When utilizing such system a somewhat rigid syntactical structure of the document is necessary. The text is composed of set tokens

10 and a set of syntactical rules to express legal relationships among the tokens. This is the de-facto approach for computer language interpreters and compilers. This approach requires a well-defined syntax and it is generally too rigid to parse free text.

The second category uses domain semantics based parsing. In this approach a parser that
15 embeds specific domain knowledge is used. Such a parser recognizes keywords and structural relationships for a well-defined domain of the document being considered. The parser is highly trained to work on a specific domain and its application to another domain requires significant changes to the parser itself.

20 Based on the above-described shortcomings, there is a need for a system that is able to automatically extract a full range of metadata from electronic documents, using a combination of text-based matching and spatial reasoning that better matches human behavior.

SUMMARY OF THE INVENTION

The present invention overcomes the deficiencies of currently available systems by using a combination of text-based matching and spatial reasoning that better matches human behavior to automatically extract a full range of metadata from electronic documents.

In one embodiment of the present invention a first processing element is configured to convert electronic documents into substantially format-invariant data files. The first processing element provides the substantially format-invariant data files to a second processing element. The second processing element is configured to receive substantially format-invariant data files, extract spatial layout facts, and provide the extracted spatial layout facts to a reasoning element. A database is configured to simultaneously provide spatial layout rules to the reasoning element; the spatial layout rules are used to extract the metadata from the substantially format-invariant data file.

Another embodiment of the present invention provides a method for automatically extracting metadata from electronic documents utilizing a first processing element and a second processing element, a reasoning element, and a database. The method includes the steps of using said first processing element to convert electronic documents to files, and using the first processing element to provide the files to the second processing element. The second processing element is utilized to receive said files and extract predetermined information. Further, the second processing element is utilized to provide extracted, predetermined information, to the reasoning element. Next, using the

database, the method provides input to the reasoning element. Using a set of rules, the reasoning element extracts metadata from the files. This extracted meta-data is provided as an output of metadata from the reasoning element.

5

BRIEF DESCRIPTION OF DRAWINGS

The accompanying drawings, which are incorporated in, and form a part of the specification, illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention.

FIG. 1 is a flowchart showing the overall architecture of one embodiment of the present invention;

FIG. 2 is a depiction of the upper portion of a scientific paper; and

FIG. 3 is a depiction of the upper portion of a scientific paper illustrating that the title is not always the first string of text on a page.

15

DETAILED DESCRIPTION

The present invention provides a method and apparatus for the extraction of metadata from electronic documents. It should be understood that this description is not intended to limit the invention. On the contrary, the invention is intended to cover alternatives, modifications and equivalents, which are included in the spirit and scope of the invention as defined by the appended claims. Furthermore, in the following detailed description of the present invention numerous specific details are set forth in order to provide a thorough understanding of the present invention. However, it will be obvious to one of

ordinary skill in the art that the present invention may be practiced without the specific details.

One embodiment of the present invention provides a spatial knowledge-based

5 methodology to document disaggregation. This approach can be easily configured to achieve improved document metadata extraction accuracy. The present embodiment is based on exploiting the visual and spatial knowledge used when reading a document. In general, within a document category, a certain visual layout can be identified for all documents within that category. For instance, a scientific paper may follow the format
10 described below. Wherein the uppercase words represent metadata in the paper and bold words denote spatial relationships and other types of relationships.

The TITLE is located on the **upper** portion of the **first page** and it is printed using the **largest font** on the **first page**;

15 AUTHORS are listed **immediately under** the TITLE in some order;

AFFILIATIONS **follow** the authors' list;

If **only one** AFFILIATION appears then **all** AUTHORS are associated with it;

The same **font** is used for all AUTHORS and, similarly, for all AFFILIATIONS;

The FIRST LEVEL HEADERS use a larger font than the SECOND LEVEL headers.

20 In the present invention, a rule-based language is used to encode the visual layout of the document. Different types of documents require different knowledge bases. A knowledge base is encoded with visual and spatial layout facts. The knowledge base

described in this embodiment deals with scientific papers appearing in conference proceedings and specialized journals. The apparatus configured to perform the steps could include a standard personal computer or other apparatus having the adequate processing power.

5

In FIG. 1 the overall architecture of the metadata extraction system is shown. The metadata extraction system retains the document's original formatting. Formatting includes both font size and text positioning on the page. Hereinafter, data that retains the original document's formatting shall be referred to as substantially format-invariant data.

10

Electronic documents **100** go to an intermediate language conversion step **102**, which is responsible for converting the electronic documents 100 into substantially format-invariant data files **104**, and capturing the spatial and visual aspects for document representation. This can generally be achieved by transferring the original document to a file from the default viewer of the document. A converted document has to undergo a spatial layout fact extraction process **106** to extract relevant spatial layout information and eliminate irrelevant information from the converted document in preparation for further processing. This is a task generally accomplished by any substantially format-invariant data printer driver or viewer.

20

One embodiment of the present invention uses a rule-based language to encode spatial facts in documents as well as rules that interpret these facts to extract metadata from

them. The rule based language output consists of a set of augmented strings of text. This additional format data is summarized in the following:

- 1) Page of the document where the specific string appears;
- 2) Absolute line counter order for each generated string;
- 3) x-y location of the lower left corner of the string bounding box in paper-dot coordinate systems;
- 4) x-y location of the upper right corner of the string bounding box in paper-dot coordinate systems;
- 5) Font metrics bounding-box extensions used to represent the given string of text.

After spatial layout facts **108** have been extracted **106** from a substantially format-invariant data file **104**, spatial layout facts **108** are subjected to spatial metadata reasoning **110**. A knowledge engineer **112** provides a set of spatial layout rules **114** that embodies the protocol for extracting the metadata **116** of interest from the provided document. A rule-based language reads the provided format-invariant data file and produces a set of spatial layout facts **108** for the rule-based language. Each fact contains information—text and spatial data—about the input substantially format-invariant data document. Rules provided by the knowledge engineer **112** reasons with the extracted facts to identify and extract relevant metadata **116** from the input documents.

The knowledge base of the present invention reasons with the spatial layout facts extracted from the substantially format-invariant data to rule-based language. The

knowledge base is encoded by means of the rules of the rule-based language. The rule set is designed to extract information from the substantially format-invariant data file such as: title, author(s), affiliation(s), mapping(s), author-affiliation, and table of contents. In this embodiment of the invention the knowledge base is comprised of 77

5 rules. The following shows the rule based language rule usage distribution for the different extraction purposes:

Extraction Purpose	Number of rules involved
Title	9
10 Author(s)	12
Affiliation(s)	10
Author Affiliation	10
Table of Contents	8
Print results	19
15 Other	9

A fundamental component of the knowledge base is the implicit fuzziness involved in the visual and spatial based metadata recognition process. For instance, with reference to the list of spatial layout fact extraction activities earlier discussed, note that:

- a. The title is not always printed on the first page using the largest font.
- 20 b. Not all papers use numbered section headers and section headers do not always use different fonts.
- c. Sometimes authors are all listed on the same line next to each other while other times the author's names are scattered across different lines.

When authors have different affiliations, different methods are employed to specify their correspondence. Two of the most popular methods are:

- i. Superscripting on author's name corresponding to the author's affiliation;
and
- 5 ii. Determining the spatial proximity of the author's name to the author's
affiliation.

Many different cases exist such as reporting affiliations as footnotes or listing authors vertically with prospective affiliations to the right on the same line.

These exceptions represent the hardest part of the artificial visual recognition

10 process. The rule-based language is coded in the knowledge base in order to be tolerant of such exceptions.

The following is an example of how the present invention extracts metadata from electronic documents. Consider the portion of a scientific paper as shown in FIG. 2.

15 Once the substantially format-invariant data to rule based language has extracted all necessary facts from the substantially format-invariant data file, the facts are processed using a rule-based language. The output of the rule based language screen for the document in FIG. 2 is as follows:

20 FILE: sigmod98

TITLE: Exploratory Mining and Pruning Optimization of Constrained
Association Rules

AUTHOR: Raymond T. Ng (1)

AUTHOR: Laks V.S. Lasshmanan (2)

AUTHOR: Jiawei Han (3)

AUTHOR: Alex Pang (1)

AFFILIATIONS 1: University of British Columbia

5 AFFILIATIONS 2: Concordia University

AFFILIATIONS 3: Simon Fraser University

---Table of Contents---

1 Introduction

3 Constrained Association Queries

10 4 Optimization Using Anti-Monotone

5 Optimization Using Succinct

6 Algorithms for Computing

6.1 Algorithms Apriori +

6.2 Algorithms Hybird (m)

15 6.3 Algorithms CAP

7 Conclusions and Future Work

The title **200** has been assembled from two lines into a single line.

The first author **202a**, the second author **202b**, the third author **202c**, and the fourth author **202d** have been correctly identified and linked to the first affiliation **204a**, the

20 second affiliation **204b**, the third affiliation **204c**, and the fourth affiliation **204d**

respectively. Notice that the system reports the first affiliation **204a** and the second affiliation **204b** "University of British Columbia" only once even though it is associated with the first author **202a** and the fourth author **202d**.

If the title **200** of a scientific document is contained in the first line of the text, or the first couple of lines of text for longer titles, a text based extraction from a substantially format-invariant data file **104** could be applied. The output data can either be displayed using a user interface, sent to a storage medium, or printed.

There are cases, as illustrated in FIG. 3 where the title is not the first string of text on the page. When information regarding the proceedings **300** of the document is above the title **302**, a straight text based approach will not be efficient in extracting the desired information.

The following rule based language was encoded with the following two hints in the knowledge base when extracting titles. Titles appear on the first page of the document and very often are printed using the largest font on the first page. Sometimes section headers use a larger, or same size, font than the title. In such a case the word “Abstract” **206** is relied on. The lines printed above “Abstract” **206** are extracted, and by using the largest font among all the lines above that word, the title can be found. The following rule based language rules are used to extract the title **200** from the paper when the word “Abstract” **206** was found on the first page as a stand-alone string:

```
(defrule CandidateTitleLines
  (declare (salience 9100) )
  (abstract-word-found ?la)
```

```

(doc (page 1) (font ?f $?)

      (absline ?n&: (< ?n ?la) ) (text ?s) )

(metrics (page 1) (font ?f) (bbh ?h1) )

=>

```

```

5      (assert (candidate-title-line ?n ?h1 ?f ?s) ) )

```

```

(defrule GetLargestFontForCandidateTitle

```

```

      (declare (salience 9090) )

      (abstract-word-found ?la)

      (candidate-title-line ?n ?h1 ?f ?)

```

```

10     (not (candidate-title-line ?

                                     ?h2&: (> ?h2 ?h1)

                                     ? ? ) )

```

```

=>

      (assert (ltf ?f) ) )

```

```

15 (defrule GetTitle1

```

```

      (declare (salience 9000) )

      (abstract-word-found ?la)

      (ltf ?f)

      (candidate-title-line ?n ?h1 ?f ?s)

```

```

20     (not (candidate-title-line

            ?n2&: (< ?h2 ?h1)

            ? ? ) )

```

```

=>

```

```

    (assert (paper-title ?n ?s)))

(defrule GetTitleNextLines

    (declare (salience 9000) )

    (abstract-word-found ?la)

5    (ltf ?f)

    (candidate-title-line ? ?hl ?f ?s)

    (not (candidate-title-line ?n2&: (< ?n2 ?n)

        ? ?f ?))

    =>

10    (assert (paper-title ?n ?s)))

(defrule GetTitleNextLines

    (declare (salience 9000))

    (abstract-word-found ?la)

    (ltf ?f)

15    ?indx <- (paper-title ?n ?s)

    (candidate-title-line ?n2&: (= (+ 1 ?n) ?n2)

        ? ?f ?t)

    =>

    (retract ?indx)

20    (bind ?s (str-cat ?s “ “ ?t) )

    (assert (paper-title ?n2 ?s) ) )

```

The first rule is CandidateTitleLines, the second rule is GetLargestFontForCandidateTitle and the third rule is GetTitleNextLines. The first rule, CandidateTitleLines, considers all

lines above the line containing the word Abstract **208** as candidates for the title **200**.

These lines include the first author **202a**, the second author **202b**, the third author **202c**, and the fourth author **202d**, and the first affiliation **204a**, the second affiliation **204b**, the third affiliation **204c**, and the fourth affiliation **204d**. At the same time the first rule,

- 5 CandidateTitleLines, extracts the font size of each text line and stores the data. In a subsequent step the rule GetLargestFontForCandidateTitle extracts the largest font from among all candidate title lines. The rule GetTitle1 gets the first line of the title **200**. The title is identified as the line having the largest font and not having any other line above it having the same size font. The last rule, GetTitleNextLines, searches for multi-line titles
- 10 and merges successive title lines having the same font type and size.

When authors' **202** names are printed using the same font as the title **200** and both titles and authors' **202** names appear above the abstract, **206**, the knowledge base may have to be further reinforced by relying on the line-position, measured along the y-coordinate.

- 15 In spatial based mapping of the first author **202a**, the second author **202b**, the third author **202c**, and the fourth author **202d**, to the first affiliation **204a**, the second affiliation **204b**, the third affiliation **204c**, and the fourth affiliation **204d**, a rule first extracts the relevant information and then attempts to match the authors with their respective affiliations **204**.

- There are many different cases to be considered since there is not necessarily a one-to-
- 20 one correlation between the authors **202** and affiliations **204**. In the simplest case, there are n authors **202** all matched to one affiliation **204**; a single rule based language takes care this type of matching. Another case arises when the number of authors **202** differs from the number of affiliations **204** and there is more than one affiliation. In such a case

a common practice, utilized by most publishers, is to use superscripts over author's **202** names and affiliations **204**. A text-based parsing protocol is exploited to resolve the associations in this case. The case now discussed is the n-to-n mapping as shown in FIG.

2. Notice that one affiliation appears twice. The first affiliation **204a** and the second

5 affiliation **204b**. In this case a spatial reasoning is operation is performed. The operation links each author **202** to that author's affiliations **204**. This is accomplished by following the rules of the rule-based language:

```
(defrule XY-AffiliationLocation
```

```
    (declare (salience 5800) )
```

```
10    (paper-affiliations ?n ?t)
```

```
    (doc (page 1) (absline ?n) (xc ?xc) (y?y) )
```

```
    =>
```

```
    (assert (xy-AFFILIATION ?n ?xc ?y) ) )
```

```
(defrule XY-AuthorLocation
```

```
15    (declare (salience 5800) )
```

```
    (paper-authors ?n ?t)
```

```
    (doc (page1) (absline ?n) (xc ?xc) (y ?y) )
```

```
    =>
```

```
    (assert (xy-author ?n ?xc ?v) ) )
```

```
20 (defrule SpatialLink-1
```

```
    Declare (salience 5800) )
```

```
    (xy-author ?n ?xp ?yp)
```

```
    (xy-affiliation ?m ?xa ?ya)
```


=>

```
(assert (link-distance ?n ?m  
  =(sqrt (+ (* (- ?ap >xa) (-?xp ?xa) )  
    (* (- ?yp ?ya) (-?yp ?ya ) ) ) ) ) ) )
```

5 (defrule SpatialLink-2

```
(declare (salience 5800) )
```

```
(n-affiliations ?n ?)
```

```
(paper-authors ?na ?t)
```

```
(not (link ?t ? ) )
```

10 (link-distance ?na ?m ?d1)

```
(paper-affiliations ?m ?tt)
```

```
(not (link-distance ?na ? ?d2&: (< ?d2 ?d1) ) )
```

=>

```
(assert (link ?t ?tt ) ) )
```

15

The rule XY-AffiliationLocation confirms the xy location, in paper dot coordinates, of the center of the string bounding box of each affiliation, i.e. the slot xc of the fact doc, which contains that location. Similarly, the rule XY-AuthorLocation confirms the bounding box center xy location of each author. In turn, the rule SpatialLink-1 computes the Euclidean distance among each possible pair author-affiliation and confirms all possible combinations using the fact link-distance. Eventually a rule, SpatialLink-2, associates each author to the spatially closest affiliation and confirms this by using the fact link.

When extracting table of contents, two basic cases are distinguished: numbered section headers and non-numbered section headers. Different sets of rules are used according to the style adopted by the paper at hand. Thus, the first thing the rule base does is

5 determine if the section headers are numbered. Section header numbering is a fundamental hint for a text-based extraction of table of contents. This is because the numbering is expected to follow a certain order throughout the paper and the numbers virtually always appear at the beginning of the line. However, headers are often not numbered, therefore an extraction based on text parsing is not applicable. In the rule
10 based system the visual properties of section headers are exploited. The section headers have a larger font than the text before and after and also have a different line-space compared to the average line-space of the entire document. Furthermore, a common header name such as "Introduction," "Overview," "Motivation," or "References" is sought in an effort to find an initial clue for the font size of the first level of headers.

15 Another embodiment of the present invention includes an apparatus for automatically extracting metadata from electronic documents. The apparatus may be an apparatus such as a conventional computer or other data processor. The apparatus includes a first processing element, a second processing element, a reasoning element, and access to a
20 database. The database may be non-local and accessed via a network, or it may be local. The first processing element is further configured to convert electronic documents into files. The first processing element is configured to provide the files to a second processing element and the second processing element is configured to

extract predetermined information from the provided file. The second processing element is further configured to provide the extracted predetermined information to the reasoning element. The database is configured to also provide input to said reasoning element. The reasoning element is configured to use a set of rules to extract metadata

5 from the files and the reasoning element provides an output of metadata. This output can go either to a printer, storage medium, or display.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225